

Weekly Report

Pingping Shang

2013.11.04~2013.11.10

本周工作

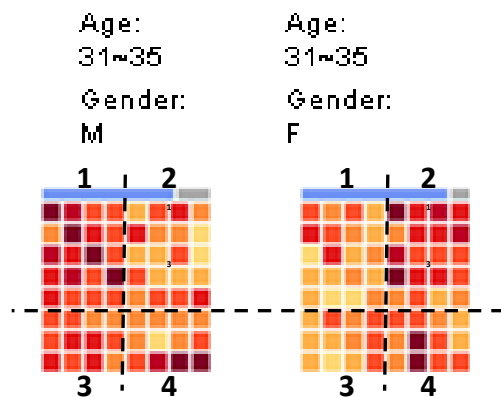
1. 之前我们的探索流程可归结为：

选择标签 => 整体展示 => 交互选择区域探索

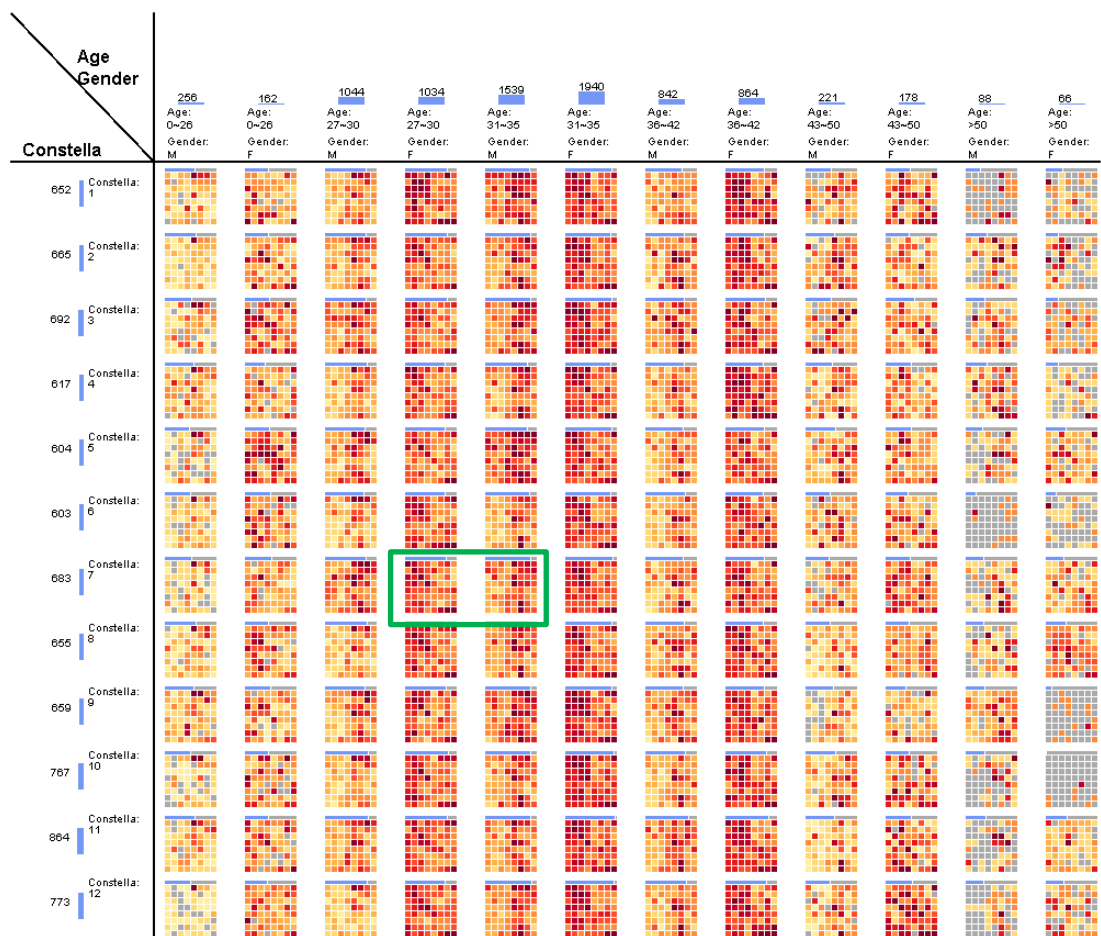
其中，交互选择区域阶段，对于选取哪块区域，系统没有自动提示，完全靠人眼选择兴趣点，但我们的结果又一直模式不明显，导致这一探索阶段也没有较大收获，为此，我们将商品进行预处理，属于同一大类的商品尽量靠近，同时将被购买次数多的商品用多个像素展示，所以这些操作都是为了提高模式的可分辨性。商品排布方式采用了在解聪在周报中提到的 **Hilbert Curve**。

将被购买次数较多的商品用多个像素表示的另一个原因：假设对于商品 A 和 B，人群 1 对商品 A、B 的购买次数都相当高，表现在像素图中，即某两个像素颜色都很深，我们很有可能得出的结论是 *人群 1 经常同时购买 A 和 B*，但由于像素只是展示的统计值，很有可能实际情况是 A、B 都是被随机购买的，只是恰巧两者的统计总数接近，A、B 之间并没有依赖性。

改变商品排布方式后得出的结果如下图（淘宝数据），其中模式的明显程度也依赖于采用的颜色归一化方式，是单个 block 归一还是全部 block 归一，我尝试了多种方式，没有比较出哪一种具有特别出色的结果。但模式确实比之前的效果集中了。举例如下两个 block(大图中绿色圈出的两个)：



1、2、3 区域的互斥现象比较容易看出



下周工作

考核报告时老师指出我的问题是思考问题不深入，不系统，不能抽象出具体问题，我应该再认真总结下在做的东西，调研相关工作，对该方向有个系统的认识，而不只是想到什么就做什么。